Ensemble Kalman Filter Analysis of Magnetic Field Models During the CHAMP-Swarm gap

Ensemble Kalman Filter Analysis of Magnetic Field Models During the CHAMP-Swarm gap

Ciarán D. Beggan[a], Kathryn A. Whaler[b]

[a] *British Geological Survey, Edinburgh, EH14 4AP, UK*
[b] *School of GeoSciences, University of Edinburgh, EH9 3LA, UK*

Abstract

Between the de-orbiting of CHAMP in September 2010 and the launch of Swarm in November 2013, there was a lack of satellite vector magnetic field data to use for main field modelling. During this period the difference between field models derived at the time and retrospective analysis (using data both before and after the vector gap) rose to around 20 nT root-mean-square (RMS). We use ensemble Kalman Filtering (EnKF) to combine steady core surface flow models of the fluid outer core with magnetic field models derived from the period when no vector satellite data were available. Since we find that the field models produced during periods without vector satellite data are just as good as the annual predictions from a flow model, there appears, at present, to be no overall benefit to using EnKF to improve field forecasting. This will remain the case until modelling can better predict secular variation.

Keywords: Ensemble Kalman filtering, magnetic field modelling, core flows, satellite vector data

1. Introduction

Since the launch of the Ørsted satellite (Olsen et al., 2000) in 1999, vector magnetic data from dedicated magnetic field missions have greatly improved models of the geomagnetic field and with it, our understanding of the behaviour

*C.D. Beggan
Email address: ciar@bgs.ac.uk (Ciarán D. Beggan)

of the various physical sources. A number of groups have produced main field models of the field generated by the internal sources (typically consisting of core, crust, (quasi-)steady ocean flow and the induced part from the ionosphere and magnetosphere) including the CHAOS (Olsen et al., 2006, 2009), GRIMM (Lesur et al., 2008, 2010) and MEME (Thomson et al., 2010; Hamilton et al., 2010) series of models. In addition, the quinquennial releases of the International Geomagnetic Reference Field (IGRF) (Finlay et al., 2010b; Thébault et al., 2015b) and World Magnetic Model (WMM) (Maus et al., 2010; Chulliat et al., 2015) benefited from the voluminous satellite dataset and the ground observatory network (Macmillan and Olsen, 2013), as well as advances in theoretical and numerical techniques.

After almost a decade in low-Earth orbit, the CHAMP mission (Reigber et al., 2002) ended in September 2010 when the satellite de-orbited at an altitude of around 290km. In November 2013, the ESA Swarm mission launched and began providing global vector data by December 2013 (Olsen et al., 2015). Thus for approximately three years, there was a 'gap' or lack of satellite vector measurements for making high-quality models of the main field. During this period, the lack of uniformly distributed global vector data led to poor spatial resolution of main field models and other problems such as the Backus effect near the magnetic equator (Backus, 1970). In addition, other effects from the spatially-biased distribution of ground magnetometer data, such as a lack of data in polar regions, arose. While the Ørsted mission, at a higher altitude of around 850 km, provided a small amount of scalar data during the intervening period, main field models had to rely on vector data solely from ground observatories. Despite these issues, several main field models in the MEME and CHAOS series were produced during the CHAMP-Swarm gap (e.g. Olsen et al., 2014). Subsequently, the flow of vector data from the Swarm mission has allowed the next generation of field models to be constructed (e.g. Finlay et al., 2016).

In this study we examine two issues. First, we assess the quality of models covering the period of the vector satellite data gap which we regard as instructive

for determining some of the errors implicit in main field models. We compare models computed at the time with later main field models that include data from both sides of the gap which employ temporal smoothing in the form of continuous splines to estimate the field where there are missing data. Secondly, we investigate whether it is possible to improve the estimate of the main field during the CHAMP-Swarm hiatus by using secular variation (SV) forecasts from core flow models. To do this we use Ensemble Kalman Filtering (EnKF) (Beggan and Whaler, 2009) to assimilate forecasts of SV from core surface flow models (Whaler and Beggan, 2015) with annual updates from a main field model generated using the data available during the hiatus.

In Section 2 we describe the contemporary field models and their differences from the 'true' field in a retrospective analysis, while in Section 3 we examine the ability of core flow models to capture main field SV. Section 4 outlines the framework for the EnKF, while Section 5 describes the results of the assimilation. We finally discuss the limitiations of modelling and assumptions in light of the results.

2. Main field modelling errors

Magnetic main field models consist of a set of time-dependent Gauss (or spherical harmonic) coefficients. Spatial values are computed from the scalar magnetic potential expanded in spherical harmonics using the Gauss coefficients up to a particular degree and order. This type of spherical harmonic representation allows the main field to be compactly described in a physically meaningful manner and allows upward and downward continuation from the Earth's surface to the magnetopause and the core-mantle boundary, respectively. Typically, the longest wavelengths of the magnetic field to degree and order 14 (around 2900 km on the Earth's surface) are captured for the core field. Recent models using Swarm data are moving toward degree and order 20 (e.g. Rother et al., 2013), though this is often in order to avoid strong spectral leakage. Above degree 14 the crustal field dominates the power spectrum.

Although the spherical harmonic representation has many advantages, one of the more obvious disadvantages is the difficulty in confidently placing errors bounds on individual coefficients (Lowes and Olsen, 2004). While it is possible to account for some of the error associated within the commissioning of the individual models (Finlay et al., 2010a) and the omission of the various sources which contribute to the measured field (Chulliat et al., 2010), there are also differences that arise between the modelling approaches of individual research groups as data selection, noise suppression and temporal and spatial damping will not be the same in each case.

2.1. Differences between DGRF/IGRF candidates

The variation between individual modelling groups can be quantified directly by examining the IGRF series of models, which are formed from a joint international effort, updated on a five-year cycle. Each final IGRF release is created from up to nine independent candidate models submitted to the IAGA Division V Modelling Working Group. The candidates are evaluated against each other and the final model (c.f. Thébault et al., 2015a). The IGRF-12 candidate models benefitted from the timely launch of the Swarm mission, so all included vector satellite as well as observatory data.

The RMS global difference ($\sqrt{dP}$) between two field models, *mod1* and *mod2*, at the Earth's surface can be calculated by (Lowes, 1966):

$$dP = \sum_{l=1}^{l_{max}} \sum_{m=0}^{l} (l+1)([(g_l^m)_{mod1} - (g_l^m)_{mod2}]^2 + [(h_l^m)_{mod1} - (h_l^m)_{mod2}]^2), \quad (1)$$

where the Gauss coefficients (g_l^m, h_l^m) to degree (l) and order (m), are arranged in a vector $\mathbf{g}$. This difference (to degree and order 13) for the Definitive Geomagnetic Reference Field (DGRF) between candidates for 2010.0 was $\sim$3 nT on average but varied from 1.7 to 6 nT. As the DGRF-2010 is a retrospective analysis of the field, this is indicative of variations in the data selection and modelling approaches of the teams.

The IGRF-2015 model is slightly different in that each team was asked to project the magnetic field ahead of time from the submission deadline in Septem-

ber to the beginning of January when the new IGRF model became effective. The variation thus is larger with the mean difference between the candidates and final model of 7.5 nT and a wider spread of 4.2 to 12.9 nT, reflecting the manner in which the field is forecast. Another factor is the end-effect on models from the use of temporal splines which are forced to reduce acceleration or some higher-order term to zero at the end knot points. With hindsight, it appears that the IGRF-2015 candidates were also affected by the emergence of a geomagnetic jerk in early 2014 which was not recognised until after the final release (Torta et al., 2015).

2.2. Differences between MEME and CHAOS models

For the period between September 2010 and November 2013, main field models generated at the time relied primarily on ground observatory and sporadic scalar field data from the Ørsted satellite at an altitude of around 850 km. The models thus suffer from a bias of data in the northern hemisphere, and a lack of vector data around the magnetic equator, as well as a globally uneven distribution in local time. The British Geological Survey (BGS) produced an annual update to MEME around the beginning of each year, using data coverage from the CHAMP era to the then-present time. There was one version of the CHAOS model produced in 2013.5 (version 4, though with occasional updates until Swarm launch). After the launch of the Swarm mission, later versions of these models were built which used the vector magnetic data either side of the gap to temporally constrain the Gauss coefficients via B-splines. Hence in these retrospective models, the magnetic field within the gap between missions has been conditioned by satellite vector data from both sides.

By comparing contemporary with retrospective models, we can investigate the expected errors which arise when there are few vector data available and the models are only well constrained in the early portions of their validity prior to September 2010. For main field models, we use five annual updates from the BGS MEME created with the magnetic global data available at the time in 2010, 2011 etc. up to 2014, which are compared to MEME2015. Note the

modelling method changed in 2015 from piecewise linear to a smooth order-6 spline representation. The MEME models are computed each year in March with the coefficients given for the start of that year. Figure 1 shows the RMS differences between MEME-201X and MEME-2015, to degree and order 14. Figure 1 also shows the CHAOS-3, -4 and -5 field models, with release dates close to 2010.0, 2013.5 and 2015.0 respectively, compared to CHAOS-6. MEME-2015 and CHAOS-5/6 use both CHAMP and Swarm data while CHAOS-3/4 and MEME-2010/11/12/13 use CHAMP and observatory vector data, and some Ørsted scalar data. MEME-2014 also uses some initial Swarm data.

Figure 1 suggests that the global model errors become larger over time, with the RMS difference by 2013 being around 20 nT. A per-coefficient analysis (not shown) suggests that degrees 1 and 5 accumulate the largest differences for the MEME models but it is degrees 1 and 2 that show the largest differences for the CHAOS models. A comparison of CHAOS-6 and MEME2015 gives an RMS difference of around 3.5 nT, on average, across the 2010-2015 period, which is in agreement with variation of DGRF-2010 candidate models, suggesting the differences arise from data selection and modelling techniques, as well as treatment of the dipole and quadrupole terms which are often contaminated by the large scale magnetospheric field, for instance. Uncertainties and differences between mathematical models also arise because of the data gaps in the polar regions, though these tend to primarily affect the zonal coefficients.

3. Forecasting with core flows

On short time-scales of less than a decade, the SV can be ascribed mainly to the advective motion of the liquid iron core carrying an embedded magnetic field (Kahle et al., 1967). Although this is incorrect for longer periods (c.f. Holme, 2007), the SV can be inverted for the advective core surface flow that captures short-term variation (Schaeffer et al., 2016).

Whaler and Beggan (2015) showed their core flow models consisting of the first two terms of a Taylor expansion of the flow with time, which we refer to

as steady flow and steady acceleration models, performed best at predicting the SV over five year periods when using a magnetic field model based on data selected from less than three years prior to the forecast. Over the past three quinquennial cycles of the IGRF and WMM series, their core flow models were better at predicting SV than the IGRF or WMM forecasts (Whaler and Beggan, 2015).

SV can be inverted for the flow expressed in its toroidal and poloidal components using the linear relationship between its spherical harmonic coefficients ($\dot{\mathbf{g}} = [\dot{g}_l^m; \dot{h}_l^m]$) and those of the toroidal and poloidal scalars of the flow. This involves the Gaunt/Elsasser matrix ($\mathbf{H}$) whose elements depend on the Gauss coefficients (Whaler, 1986). The main field, SV and flow coefficient expansions are truncated at degree and order $l = 14$. It is also possible to incorporate magnetic secular acceleration (SA) ($\ddot{\mathbf{g}}$) into the inversion to allow estimation of flow acceleration. In this study, acceleration is included up to degree and order $l = 8$.

Following the approach of Whaler and Beggan (2015), we used two different sets of magnetic field data covering the period 2000–2010 to generate SV and SA estimates. The first was vector monthly mean values based on night-time data from up to 160 global magnetic observatories. As this network is very unevenly spatially distributed, we also used satellite data to provide global coverage. We calculated 'virtual observatory' (VO) (Mandea and Olsen, 2006) monthly field component time series from CHAMP vector data (version 51) on a grid of 648 points at equal latitude and longitude spacings of 10° in colatitude and longitude. Each VO was located at a nominal altitude of 400 km and encompassed satellite data within a 400 km radius from the centre point.

Annual first differences of main field and SV values provided SV and SA estimates, respectively, at both ground observatories and VOs. The difference between month n+12 and month n was designated to be the value at month n+6, giving time series of SV and SA. From the VO method, the variance of each monthly solution for the individual magnetic field components (and hence the variances of the SV and SA values derived from them) can be computed.

The uncertainties of the ground observatory SV and SA data are unknown but assumed to be small (c.f. Lesur et al., 2017). We assigned them arbitrary values of 1 nT/yr and 1 nT/yr^2 in each component.

We seek flows ($\hat{\mathbf{m}}$) which can be obtained from the SV and SA using the standard L_2 least-squares minimisation norm (Gubbins, 1983). We then apply an additional step using an iterative L_1 norm minimisation technique as described by Beggan and Whaler (2008). This will account for an incorrect guess of the observatory SV and SA uncertainties, and improves the fit to the data. The magnetic field SV and SA have been inverted using two types of assumption about the flow:

- SF SV only: using magnetic SV data for a steady flow only
- SF/SA: using magnetic SV and SA data to invert for both a steady flow and steady acceleration

We also inverted SV and SA data from different periods and lengths of time: 2001–2010, 2001-2007, 2005–2010 and 2007–2010. The main features of the steady part of the flow are common to all the models, but the acceleration changes markedly depending on how many and which years are included in the inversion.

To forecast the change of the magnetic field, Gauss coefficients from the CHAOS-6 model for 2010.0 were used as the starting field model. The field was advected forward on a monthly timestep (k) for five years using the equation:

$$\mathbf{g}_{k+1} = \mathbf{g}_k + (\mathbf{H}_k \hat{\mathbf{m}})/12 \tag{2}$$

where the Gaunt/Elsasser matrix, $\mathbf{H}_k$, is updated at every timestep using the main field coefficients forecast from the previous time step. To evaluate the validity of this forecast, in Figure 2, we show the forecasts of magnetic field change from 2010.0 using different core flow models, relative to CHAOS-6. The different colours in the figure show the span of magnetic field data used for each flow. The upper panel shows the steady flows inverted with SV only and the lower panel gives the steady flow with steady acceleration (labelled SF/SA).

By inspection, it can be seen that the prediction based on the SF/SA flow using data spanning 2007–2010 gives the lowest RMS difference from CHAOS-6 at 2015.0 of 73 nT. This compares to a total SV change over the 2010–2015 period of 440 nT, and thus the flow captures over 83% of the variation. Table 1 gives the numerical differences for the models in Figure 2, as well as the IGRF and WMM values. We conclude that the best forecast model is derived from satellite and ground magnetic SV and SA data spanning 2007–2010.

4. Ensemble Kalman Filtering

Ensemble Kalman Filtering (EnKF) is a method for optimally combining models of observational information with a physical model of the process using the statistical representation of their associated uncertainties (c.f. Evensen, 2003). It is used extensively to improve the accuracy of weather forecasts by exploring the sensitivity of systems to minor perturbations or initial conditions. Its use in data assimilation within geomagnetism has increased over the past decade, particularly for forcing geodynamo models to behave in a more Earth-like manner (e.g. Aubert, 2013; Tangborn and Kuang, 2015; Barrois et al., 2017).

In EnKF, the state of a dynamic process at any particular time can be represented as a vector in n-dimensional space, where n is the number of parameters in the system. The uncertainty of the process is represented by perturbing the inputs randomly by a known variance (with zero-mean) to produce an ensemble of states. The evolution of the states through time is controlled by propagating the ensemble forward using model equations of the system behaviour. When an observation is available, it can be optimally assimilated into the ensemble by applying the standard Kalman Filter equations. With a sufficiently large ensemble, the mean state represents the most likely value for the process at the time. The evolution of the ensemble can be explored by examining the spread of the states about the mean.

A traditional Kalman Filter is implemented in two steps: (1) prediction of the evolution of the model state by dynamic equations believed to represent

the system adequately and (2) assimilation of a measurement to correct any accumulated error in the model. At time k, the optimal blending of a forecast state ($\mathbf{x}_k^f$) and measurement ($\mathbf{z}_k$) to generate the assimilated state vector, $\mathbf{x}_k^a$, is through the so-called Kalman gain matrix ($\mathbf{K}_k$):

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k(\mathbf{z}_k - \mathbf{x}_k^f) \tag{3}$$

with

$$\mathbf{K}_k = \mathbf{P}_k^f(\mathbf{P}_k^f + \mathbf{Q})^{-1} \tag{4}$$

where $\mathbf{P}_k^f$ is the covariance of the model and $\mathbf{Q}$ is the covariance of the data measurement. The balance between the model (in this case our $\mathbf{g}_k$) and measurement error controls the assimilation step and it is these values that we seek to extract from the analyses in the previous two sections.

We follow the methodology of Beggan and Whaler (2009) with an EnKF of 1000 members, progressing in two stages: (i) a forecast step based upon the flow model and (ii) an assimilation step to infuse the coefficients of a contemporary field model into the system in order to update and correct the trajectory. We use the flow models inverted from the 2007-2010 magnetic field data to drive the EnKF in forecast mode for one year (Eq. 2) and the Gauss coefficients of the MEME-201X field models computed from the data available at the time for the annual assimilation (Eq. 3). We compare the results to the forecast from the 2007-2010 model to see if there are improvements beyond this.

To initialise the system, we start at 2009.0 and specify the (assumed diagonal) covariance matrix elements as follows for the $\mathbf{P}$ and $\mathbf{Q}$ matrices. To estimate the model error ($\mathbf{P}$) generated by a flow model, we use the per coefficient differences (in nT/yr) between a flow forecast and the true SV field coefficients after one year. For the magnetic field (measurement error, $\mathbf{Q}$), the differences are between the MEME2014 and MEME2015 field models at 2014.0. These errors are shown in Figure 3, with degree 1 showing the largest difference.

The forecast (prediction) of the field is driven forwards by the summation of the field coefficients and the monthly SV from the flow model which is perturbed by a random matrix with zero mean and standard deviation computed from the

variance of the flow over time. In addition, at each time step, model noise is added to simulate the variance of the ensemble, forcing it to grow at each forecast iteration. The model noise is controlled by the size of the time-step (Δt) from the flow model (Eq. 2) which is one month, the standard deviation of the SV from the previous iteration, and a fixed parameter ($\rho = 0.009$) which can be used to control the time correlation of the noise, as required (Evensen, 2003).

As time progresses, the forecast field model will begin to diverge from the actual field. Measured (or modelled) values can be input into the ensemble to update (and correct) it. The measured data have associated errors which are used to generate a perturbed ensemble of measurements, whose mean is equal to the input data. The perturbed measurements are assimilated into the overall ensemble using the Kalman Filter algorithm. The forecast process is repeated each month until a measurement becomes available for assimilation into the ensemble. By changing the weighting of $\mathbf{P}$ and $\mathbf{Q}$ in the EnKF we can investigate strategies for improving the overall forecast of the SV.

In our first experiment, we assume that the errors models have the weighting ascribed in Figure 3. In this case, the flow model errors are smaller than the field model and so the forecast state ($\mathbf{x}^f$) is more highly weighted in the assimilation step in Equation (3). Figure 4 shows the outcome of this ensemble forecast. The RMS differences are with respect to MEME-2015 magnetic field model to degree 14. Note that at the time of each measurement assimilation, the ensemble collapses back to this point.

Figure 4 shows the RMS differences for the EnKF forecast using a SF/SA model. The individual members are in green (1000 of them), the ensemble mean is in black with the $\pm 1\sigma$ values of the ensemble in red. The gray line is the forecast from the 2007–2010 model (light blue line shown in Figure 2 (lower panel)) that produced the best forecast of the flow models tested. As can be seen, compared to the best performing simple forecast flow model, there is not much improvement in the forecasting ability; after five years, the reduction is less than 3 nT. This suggests that there is little benefit in assimilating a field

model at the beginning of each year when it does not have much weight in the Kalman gain matrix (**K**).

For the second experiment, we assumed that the errors for the field model are too pessimistic. In this case, we simply divided the individual measurement errors by some factor (e.g. 5, 10, 20, 50 or 100) to make them progressively smaller. The **Q** is now tiny in Equation (4) and so the measurement (**z**) is essentially error-free. Figure 5 shows the outcome of this assumption for a factor of 50 (again for the SF/SA flow model).

The forecast driven by the flow model drifts further away from the true field over the year but responds strongly when the field models are assimilated each year. For the first two years, the assimilation step improves the forecast, but in 2012 it actually makes it worse. This may be related to the geomagnetic jerk in 2011 (Chulliat and Maus, 2014) which was not well captured within the then-contemporary field model. Following this, assimilation starts to improve the forecast again, and in 2015 it reduces the RMS error sharply to values of around 5 nT. The RMS difference remains below 31 nT throughout the period of 2010-2015. This is better than the result from using the flow model forecast of Figure 2.

Figure 6 illustrates the spatial effect of the field assimilation into the forecast. The spatial differences in each of the three magnetic field components are shown just before and after an assimilation, which occurs at the start of each year. The first assimilation in 2010.0 has a slight impact on the differences, but by 2015.0 the assimilation of the field model produces a strong reduction, particularly in the Z component. This suggests that the flow model is not fully capturing the changes in the Southern hemisphere compared to the Northern hemisphere. The larger field differences in the Southern hemisphere may also reflect the data distribution within the main field model as fewer observatories exist in this region.

In another experiment (not shown) the flow model error was divided by 50 to make **P** dominant in Eq. (4). In this scenario, the forecast then tends toward the simple flow forecast (gray line). As noted we varied the amount by which

we reduced the field model errors (e.g. 5, 10, 20, 100) but found that above 50 (for which the results are shown in Figure 5) there was no discernible change in the forecasts.

Finally, we examined the use of an error model based on the expected covariances of the main field coefficients themselves. Based on a suggestion by one of the reviewers, we used equation (6.1) of Lowes and Olsen (2004) to build a relative covariance matrix to represent realistic correlation between the Gauss coefficients of a satellite-derived field model. We use the unscaled Lowes and Olsen correction factor for each Gauss coefficient $\sigma_{m,n} = 0.27 + (1.81 + 13.18/n)\exp[(nm)/4.49] + (1.62 + 9.83/n)\exp[m/1.09]$ to create a covariance matrix for the field model. Note, their equation was based on an analysis of an Ørsted main field model (OSFM4), which had a dense coverage of vector data. Hence, we should expect errors in these model coefficients to be smaller than to the MEME201X models during 2011-2014.

The full covariance matrix enters the ensemble via the $\mathbf{Q}$ matrix where it acts to simulate the measurement error (see equation (13) of Beggan and Whaler (2009)). However, the magnitude of the estimated variances of the Gauss coefficients in Lowes and Olsen (2004) are relatively small. They suggest that Gauss coefficients up to degree 14 have variances less than 10^{-2} nT^2, which is well below the level of variation we used in this study. Experiments with the more realistic covariance matrix (not shown) suggested it has little to no effect on the overall performance of the ensemble, as the relative size of random numbers in the present ensemble are two or more orders of magnitude larger. This suggests we are already being pessimistic about the variance of the model coefficients in any case, even if we do randomly allocate them.

5. Discussion

The aim of this study was two-fold. Firstly, we looked at the use of core flow forecasts to improve field modelling during gaps in vector data from satellite missions. The second aim was to determine the best balance between the errors

assigned to the flow and field to produce an optimal forecast with the benefit of retrospective field models available for re-analysis.

In general the flow model type and length of magnetic field data inverted to create the flow, strongly influences the fidelity of the forecast. Whaler and Beggan (2015) showed that a hindcast flow can recreate the magnetic field over the time era it covers to within 10 nT over 5 years. However large-scale steady flows do not capture more rapid dynamic changes such as jerks and presently ignore the effects of diffusion. They also do not contain small scales (above degree 14), which Barrois et al. (2017) point out are important in fully describing the field change even over five year periods.

We also note that the use of relatively simple covariance matrices in the ensemble calculations implies the matrices P and Q are not fully exploited. Though we examined a realistic covariance model from Lowes and Olsen (2004), it was found that introducing additional complexity, for example to compensate for limitations due to polar gaps, did not alter the results significantly. Although other methods can be envisaged for choosing more realistic covariance matrices e.g. that reflect the unequal distribution of observatories in the individual years of the MEM201X models, it is unlikely they would have a strong influence on the overall forecast accuracy.

By using the EnKF to combine forecasts from a SF/SA core flow models covering 2007–2010 with those from main field models built without vector satellite data during the CHAMP-Swarm gap, we attempted to deduce the optimal balance between realistic flow and field model errors. The RMS differences in the forecasts in Figure 4 show that, if we ascribe equal weight to the error models of the flow and main field (Figure 3), the forecasts are similar to using the flows by themselves.

On the other hand, if we essentially assume the field model predictions are error-free in the assimilation step (Figure 5), the forecast is much better. However, we can only reduce the RMS difference to that of the field model itself (c.f. Figure 1). In some circumstances, assimilating a contemporary field model can make the forecast worse, as in 2012. Hence, we can only do as well as the

'better' part of models in the EnKF system. Given that the field models produced during periods of no-vector satellite data are just as good as the annual predictions from a flow model, there appears at present to be no overall benefit to using EnKF. We suggest that this will remain the case until flow inversion (or geodynamo) models can predict the SV better (c.f. Baerenzung et al., 2016).

6. Conclusions

We examined the use of core flow forecasts to improve field modelling during periods where vector magnetic data from satellite missions were unavailable. We sought to determine the best balance between the errors assigned to the flow and field models in order to produce an optimal forecast of the magnetic field using a Ensemble Kalman Filter.

We find that by assuming the field models are error-free in the ensemble assimilation the forecast is much better than using realistic errors from a flow model. However, we can only improve the foreast performance to the 'better' part of models used in the EnKF. Hence the overall forecast of field change is not significantly improved by using an EnKF approach. At present, there appears to be no strong benefit to using EnKF in this manner. We suggest that this will remain the case until flow models can better predict the secular variation of the magnetic field.

7. Acknowledgements

The results presented in this paper rely on data collected at magnetic observatories. We thank the national institutes that support them and INTERMAGNET for promoting high standards of magnetic observatory practice (www.intermagnet.org). We also acknowledge the use of data from the Ørsted, CHAMP and Swarm satellite missions. We thank the two anonymous reviewers for their constructive comments and suggestions on an earlier draft of the manuscript. In particlar it was suggested that we examine the correlation between main field coefficients using Lowes and Olsen (2004). This research did not receive any specific grant

from funding agencies in the public, commercial, or not-for-profit sectors. This manuscript has been approved by the Executive Director of the British Geological Survey (NERC).

References

Aubert, J., 2013. Flow throughout the Earth's core inverted from geomagnetic observations and numerical dynamo models. Geophysical Journal International 192, 537–556. doi:10.1093/gji/ggs051, arXiv:http://gji.oxfordjournals.org/content/192/2/537.full.pdf+html.

Backus, G.E., 1970. Nonuniqueness of the external geomagnetic field determined by surface intersity measurements. J. Geophys. Res. 75, 6339–6341.

Baerenzung, J., Holschneider, M., Lesur, V., 2016. The flow at the Earth's coremantle boundary under weak prior constraints. J. Geophys. Res. Solid Earth 121, 13431364. doi:10.1002/2015JB012464.

Barrois, O., Gillet, N., Aubert, J., 2017. Contributions to the geomagnetic secular variation from a reanalysis of core surface dynamics. Geophysical Journal International 211, 50–68. doi:10.1093/gji/ggx280.

Beggan, C., Whaler, K., 2008. Core flow modelling assumptions. Phys. Earth Planet. Int. 167, 217–222.

Beggan, C., Whaler, K., 2009. Forecasting change of the magnetic field using core surface flows and ensemble kalman filtering. Geophysical Research Letters 36, L18303. doi:10.1029/2009GL039927.

Chulliat, A., Macmillan, S., Alken, P., Beggan, C., Nair, M., Hamilton, B., Woods, A., Ridley, V., Maus, S., Thomson, A., 2015. The US/UK World Magnetic Model for 2015-2020. Technical Report. National Geophysical Data Center, NOAA. doi:10.7289/V5TB14V7.

Chulliat, A., Maus, S., 2014. Geomagnetic secular acceleration, jerks, and a localized standing wave at the core surface from 2000 to 2010. Journal of Geophysical Research: Solid Earth 119, 1531–1543. doi:10.1002/2013JB010604.

Chulliat, A., Thebault, E., Hulot, G., 2010. Core field acceleration pulse as a common cause of the 2003 and 2007 geomagnetic jerks. Geophysical Research Letters 37, L07301. doi:10.1029/2009GL042019.

Evensen, G., 2003. The Ensemble Kalman Filter: theoretical formulation and practical implementation. Ocean Dynamics 53, 343367. doi:10.1007/s10236-003-0036-9.

Finlay, C., Maus, S., Beggan, C., Hamoudi, M., Lowes, F., Olsen, N., Thébault, E., 2010a. Evaluation of candidate geomagnetic field models for IGRF-11. Earth Planets and Space 62, 787804. doi:10.5047/eps.2010.11.005.

Finlay, C.C., Maus, S., Beggan, C.D., Bondar, T.N., Chambodut, A., Chernova, T.A., Chulliat, A., Golovkov, V.P., Hamilton, B., Hamoudi, M., Holme, R., Hulot, G., Kuang, W., Langlais, B., Lesur, V., Lowes, F.J., Lühr, H., Macmillan, S., Mandea, M., McLean, S., Manoj, C., Menvielle, M., Michaelis, I., Olsen, N., Rauberg, J., Rother, M., Sabaka, T.J., Tangborn, A., Tøffner-Clausen, L., Thébault, E., Thomson, A.W.P., Wardinski, I., Wei, Z., Zvereva, T.I., 2010b. International Geomagnetic Reference Field: the eleventh generation. Geophysical Journal International 183, 1216–1230. doi:10.1111/j.1365-246X.2010.04804.x, arXiv:http://gji.oxfordjournals.org/content/183/3/1216.full.pdf+html.

Finlay, C.C., Olsen, N., Kotsiaros, S., Gillet, N., Tøffner-Clausen, L., 2016. Recent geomagnetic secular variation from swarm and ground observatories as estimated in the CHAOS-6 geomagnetic field model. Earth, Planets and Space 68, 112. doi:10.1186/s40623-016-0486-1.

Gubbins, D., 1983. Geomagnetic field analysis - I. Stochastic inversion. Geophys. J. R. Astr. Soc. 73, 641–652.

Hamilton, B., Macmillan, S., Thomson, A., 2010. The BGS magnetic field candidate models for the 11th generation IGRF. Earth Planets and Space 62, 737. doi:10.5047/eps.2010.05.005.

Holme, R., 2007. Treatise on Geophysics. Elsevier. volume 8. chapter Large Scale Flow in the Core. pp. 107–130.

Kahle, A., Vestine, E., Ball, R., 1967. Estimated surface motions of the Earth's core. J. Geophys. Res. 72, 1095–1108.

Lesur, V., Heumez, B., Telali, A., Lalanne, X., Soloviev, A., 2017. Estimating error statistics for Chambon-la-forêt observatory definitive data. Annales Geophysicae 35, 939–952. doi:10.5194/angeo-35-939-2017.

Lesur, V., Wardinski, I., Hamoudi, M., Rother, M., 2010. The second generation of the GFZ Reference Internal Magnetic Model: GRIMM-2. Earth, Planets and Space 62, 6. URL: https://doi.org/10.5047/eps.2010.07.007, doi:10.5047/eps.2010.07.007.

Lesur, V., Wardinski, I., Rother, M., Mandea, M., 2008. GRIMM: the GFZ Reference Internal Magnetic Model based on vector satellite and observatory data. Geophys. J. Int. 173, 382–394. doi:10.1111/j.1365-246X.2008.03724.x.

Lowes, F., 1966. Mean-square values on a sphere of spherical harmonic vector fields. J. Geophys. Res. 71, 2179.

Lowes, F.J., Olsen, N., 2004. A more realistic estimate of the variances and systematic errors in spherical harmonic geomagnetic field models. Geophys. J. Int. 157, 10271044. doi:10.1111/j.1365-246X.2004.02256.x.

Macmillan, S., Olsen, N., 2013. Observatory data and the Swarm mission. Earth, Planets and Space 65, 15. URL: https://doi.org/10.5047/eps.2013.07.011, doi:10.5047/eps.2013.07.011.

Mandea, M., Olsen, N., 2006. A new approach to directly determine the secular variation from magnetic satellite observations. Geophys. Res. Lett. 33, L15306. doi:10.1029/2006GL026616.

Maus, S., Macmillan, S., McLean, S., Hamilton, B., Thomson, A., Nair, M., Rollins, C., 2010. The US/UK World Magnetic Model for 2010-2015. Technical Report. NOAA Technical Report NESDIS/NGDC.

Olsen, N., Holme, R., Hulot, G., Sabaka, T., Neubert, T., Tøffner-Clausen, L., Primdahl, F., Jorgensen, J., Leger, J.M., Barraclough, D., Bloxham, J., Cain, J., Constable, C., Golovkov, V., Jackson, A., Kotze, P., Langlais, B., Macmillan, S., Mandea, M., Thomson, A., Voorhies, C., 2000. Oersted initial field model. Geophy. Res. Lett. 27, 3607–3610.

Olsen, N., Hulot, G., Lesur, V., Finlay, C.C., Beggan, C., Chulliat, A., Sabaka, T.J., Floberghagen, R., Friis-Christensen, E., Haagmans, R., Kotsiaros, S., Lhr, H., Tffner-Clausen, L., Vigneron, P., 2015. The Swarm Initial Field Model for the 2014 geomagnetic field. Geophysical Research Letters 42, 1092–1098. URL: http://dx.doi.org/10.1002/2014GL062659, doi:10.1002/2014GL062659. 2014GL062659.

Olsen, N., Lühr, H., Finlay, C.C., Sabaka, T.J., Michaelis, I., Rauberg, J., Tøffner-Clausen, L., 2014. The CHAOS-4 geomagnetic field model. Geophys. J. Int. 197, 815–827. doi:10.1093/gji/ggu033.

Olsen, N., Lühr, H., Sabaka, T., Mandea, M., Rother, M., Tøffner-Clausen, L., 2006. CHAOS: a model of the Earth's magnetic field derived from CHAMP, Oersted, and SAC-C magnetic satellite data. Geophys. J. Int. 166, 67–75.

Olsen, N., Mandea, M., Sabaka, T.J., Tøffner-Clausen, L., 2009. CHAOS-2a geomagnetic field model derived from one decade of continuous satellite data. Geophy. J. Int. 179, 1477–1487. doi:10.1111/j.1365-246X.2009.04386.x.

Reigber, C., Lühr, H., Schwintzer, P., 2002. CHAMP mission status. Adv. Space Res. 30, 129–134. doi:10.1016/S0273-1177(02)00276-4.

Rother, M., Lesur, V., Schachtschneider, R., 2013. An algorithm for deriving core magnetic field models from the Swarm data set. Earth, Planets and Space 65, 3. URL: https://doi.org/10.5047/eps.2013.07.005, doi:10.5047/eps.2013.07.005.

Schaeffer, N., Silva, E.L., Pais, M., 2016. Can core flows inferred from geomagnetic field models explain the Earth's dynamo? Geophysical Journal International 204, 868–877. doi:10.1093/gji/ggv488.

Tangborn, A., Kuang, W., 2015. Geodynamo model and error parameter estimation using geomagnetic data assimilation. Geophysical Journal International 200, 664–675. doi:10.1093/gji/ggu409.

Thébault, E., Finlay, C.C., Alken, P., Beggan, C.D., Canet, E., Chulliat, A., Langlais, B., Lesur, V., Lowes, F.J., Manoj, C., Rother, M., Schachtschneider, R., 2015a. Evaluation of candidate geomagnetic field models for igrf-12. Earth, Planets and Space 67, 112. URL: https://doi.org/10.1186/s40623-015-0273-4, doi:10.1186/s40623-015-0273-4.

Thébault, E., Finlay, C.C., Beggan, C.D., Alken, P., Aubert, J., Barrois, O., Bertrand, F., Bondar, T., Boness, A., Brocco, L., Canet, E., Chambodut, A., Chulliat, A., Coïsson, P., Civet, F., Du, A., Fournier, A., Fratter, I., Gillet, N., Hamilton, B., Hamoudi, M., Hulot, G., Jager, T., Korte, M., Kuang, W., Lalanne, X., Langlais, B., Léger, J.M., Lesur, V., Lowes, F.J., Macmillan, S., Mandea, M., Manoj, C., Maus, S., Olsen, N., Petrov, V., Ridley, V., Rother, M., Sabaka, T.J., Saturnino, D., Schachtschneider, R., Sirol, O., Tangborn, A., Thomson, A., Tøffner-Clausen, L., Vigneron, P., Wardinski, I., Zvereva, T., 2015b. International geomagnetic reference field: the 12th generation. Earth, Planets and Space 67, 79. doi:10.1186/s40623-015-0228-9.

Thomson, A.W.P., Hamilton, B., Macmillan, S., Reay, S.J., 2010. A novel weighting method for satellite magnetic data and a new global magnetic field model. Geophysical Journal International 181, 250–260. doi:10.1111/j.1365-246X.2010.04510.x.

Table 1: RMS differences relative to CHAOS-6 from different core flow model forecasts for 2010.0 to 2015.0. For comparison the differences for IGRF-11 and WMM2010 SV forecasts at 2015.0 are 82.3 and 102.4 nT, respectively. Units: nT.

Data span	SV only	SF/SA
2001-2007	111.7	110.6
2001-2010	82.9	104.8
2005-2010	77.9	82.2
2007-2010	74.8	73.0

Torta, J.M., Pavón-Carrasco, F.J., Marsal, S., Finlay, C.C., 2015. Evidence for a new geomagnetic jerk in 2014. Geophysical Research Letters 42, 7933–7940. URL: http://dx.doi.org/10.1002/2015GL065501, doi:10.1002/2015GL065501. 2015GL065501.

Whaler, K.A., 1986. Geomagnetic evidence for fluid upwelling at the core-mantle boundary. Geophys. J. R. Astr. Soc. 86, 563–588.

Whaler, K.A., Beggan, C.D., 2015. Derivation and use of core surface flows for forecasting secular variation. Journal of Geophysical Research: Solid Earth 120, 1400–1414. URL: http://dx.doi.org/10.1002/2014JB011697, doi:10.1002/2014JB011697. 2014JB011697.

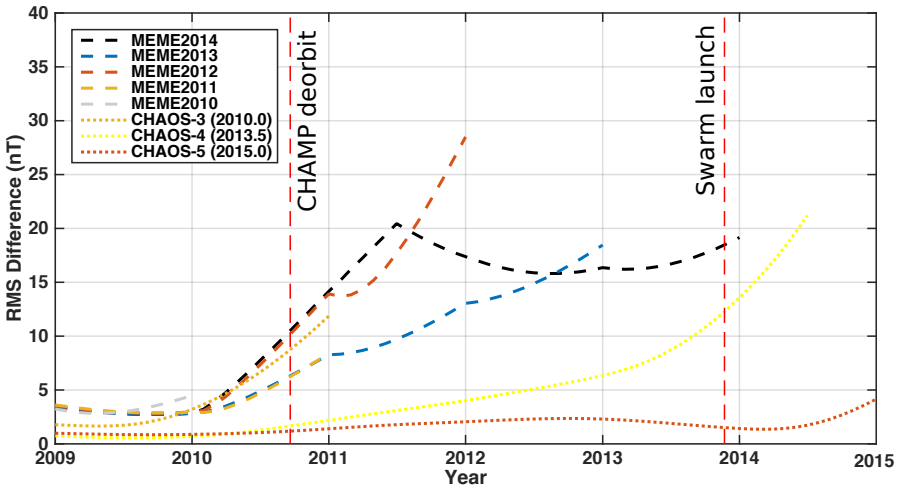

Figure 1: Comparison of root-mean-square (RMS) differences (in nT) of MEME-201X with MEME-2015 and CHAOS-X with CHAOS-6. Differences are to degree and order 14. Model release dates are shown in the legend.

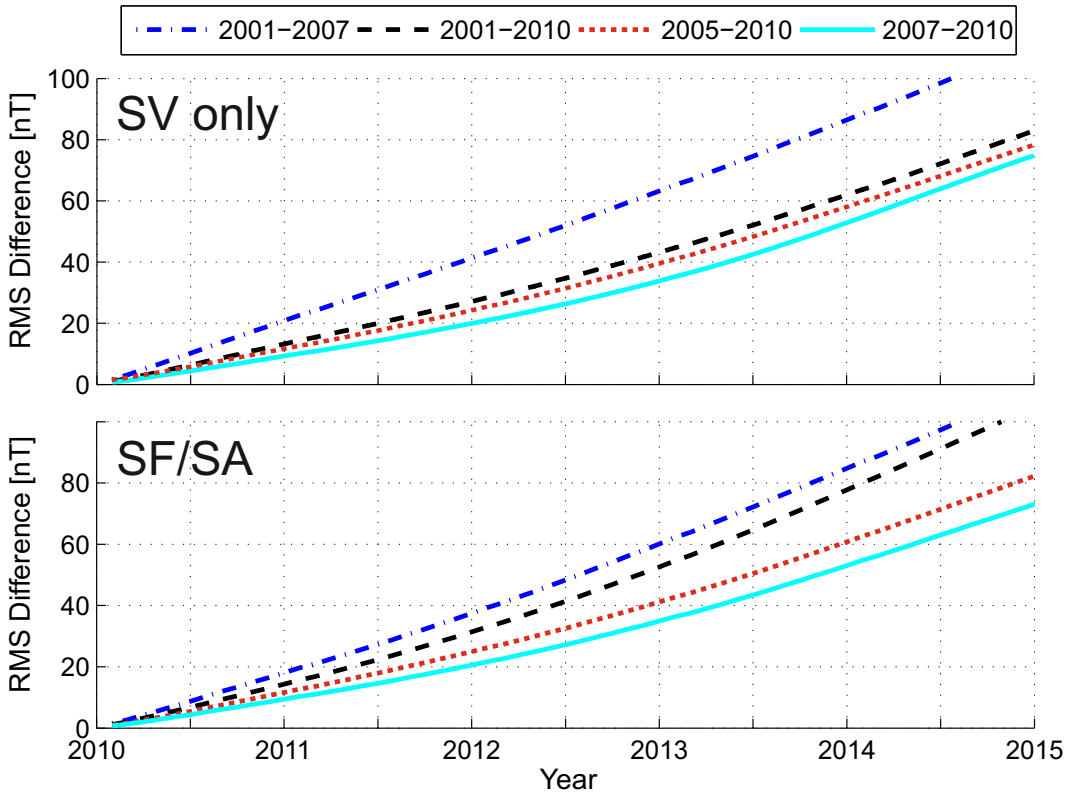

Figure 2: Root-mean-square (RMS) differences over 2010–2015 between CHAOS-6 and predictions based on core surface flow models derived assuming: (upper) SV magnetic data only with no flow acceleration; (lower) SV and SA magnetic data and including flow acceleration. Different time periods of magnetic data prior to the forecast are used to compute the flows.

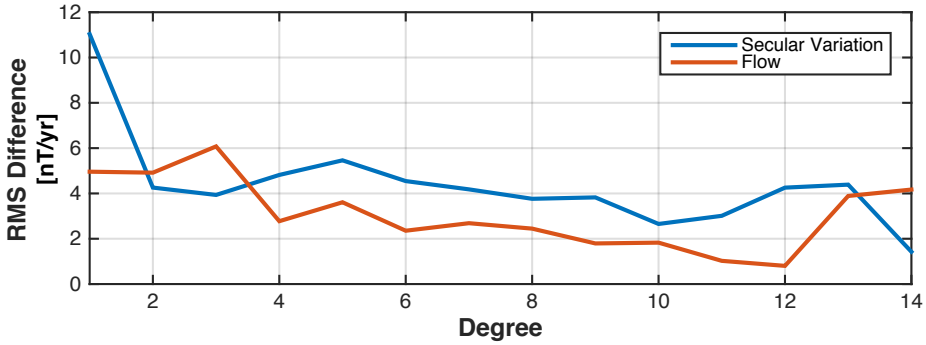

Figure 3: Root-mean-square (RMS) differences per degree of the secular variation between the MEME2014 and MEME2015 field models for 2014–2015 (blue line) and the RMS difference between a flow model forecast and the true SV field (red line) after one year. See text for details.

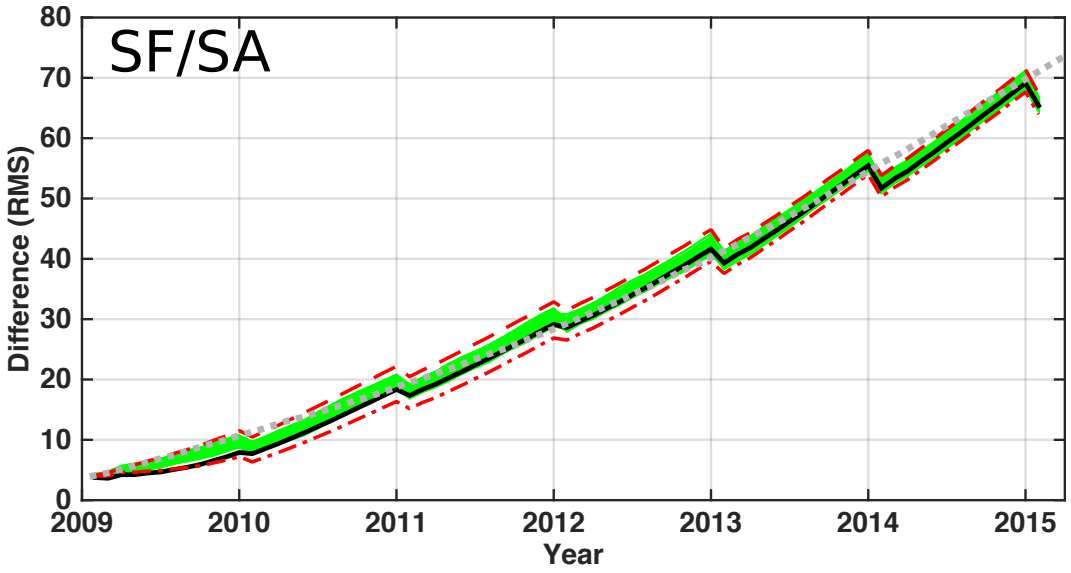

Figure 4: Comparison of root-mean-square (RMS) differences (in nT) over 2009-2015 assuming that the errors of the flow model and field model have equal weighting (based on their assumed uncertainties) in the EnKF assimilation step. Individual members are in green, the ensemble mean is in black with the $\pm 1\sigma$ of the ensemble in red. The gray line is the forecast from the 2007–2010 model shown in Figure 2. Differences are relative to MEME2015 to degree and order 14.

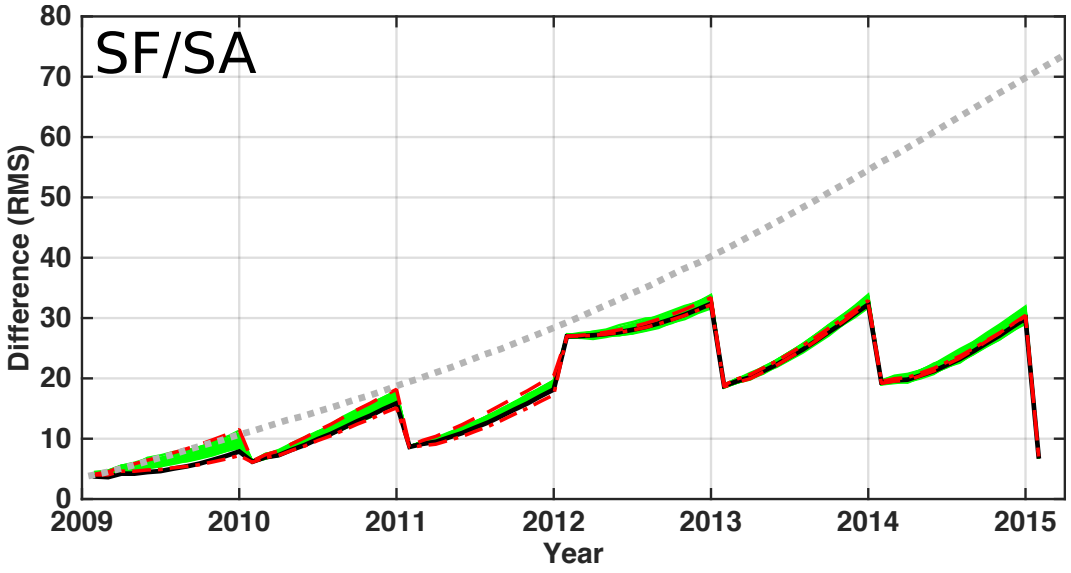

Figure 5: Comparison of root-mean-square (RMS) differences (in nT) over 2009-2015 assuming that the field model errors are reduced by a factor 50 in the EnKF assimilation step. The individual members are in green, the ensemble mean is in black with the $\pm 1\sigma$ of the ensemble in red. The gray line is the forecast from the 2007–2010 model shown in Figure 2. Differences are relative to MEME2015 to degree and order 14.

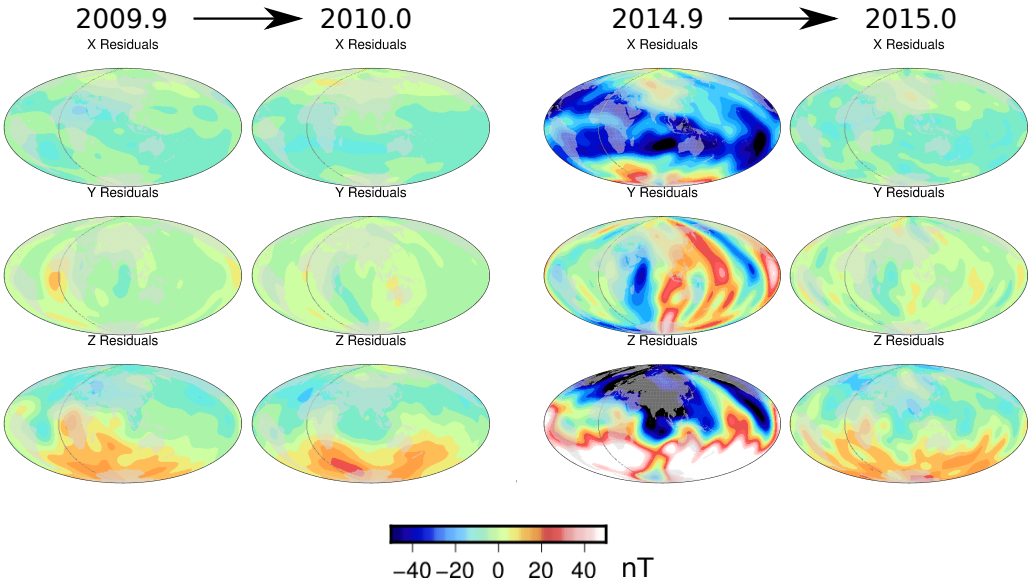

Figure 6: Differences in the X, Y and Z components between the forecast model prior to and after assimilation of the then-available main field model. Left panels: The first assimilation in 2010.0; Right panels: The final assimilation in 2015.0. Central meridian is 90°E. The Greenwich meridian is shown as a black line.